6.0 THE ANALYSIS OF VARIANCE

6.1 One-way Analysis of Variance

We will begin with the simplest case, the one-way analysis of k sets of observations (the t-test considers k=2). The ANOVA is usually presented as a table showing the sums of squares, degrees of freedom, mean squares and the associated F-tests of significance (named for the pioneer worker and originator of the test, R. A. Fisher, who devised much of the methodology in the 1920's and 1930's).

The basic calculations depend only on algebraic identities yielding the Sums of Squares. These hold for any k sets of numbers so there are no assumptions involved in the basic calculations. We bring in various assumptions in order to develop statistical tests of significance.

We will subject one set of data (the pheasant count data of Table 6.1) to several different forms of ANOVA to demonstrate the mechanics of calculations for various arrangements of data. The assumptions involved in F-tests will be discussed more fully later, after we examine the basic calculations.

Table 6.1. Pheasant call count data reported by S.M. Carney and G. A. Petrides Journal of Wildlife Management 21:393, 1957

	OBSERVERS					
STATIONS	A	B	C	D	E	F
1	39	33	33	32	29	27
2	46	36	32	30	35	35
3	45	36	44	31	31	23
4	15	25	29	18	18	14
5	17	14	14	9	14	7
6	27	24	26	14	20	15
7	24	19	15	13	19	15
8	22	22	22	13	16	13
9	28	35	33	32	26	28
10	26	24	23	26	22	17
11	12	13	5	9	8	8
12	8	11	9	9	12	7
13	6	5	9	4	10	3
14	7	6	7	2	7	6
15	7	11	11	6	10	6
16	9	11	9	6	10	4
17	1	2	4	2	5	0
18	5	4	4	4	6	2
19	3	2	3	1	3	1

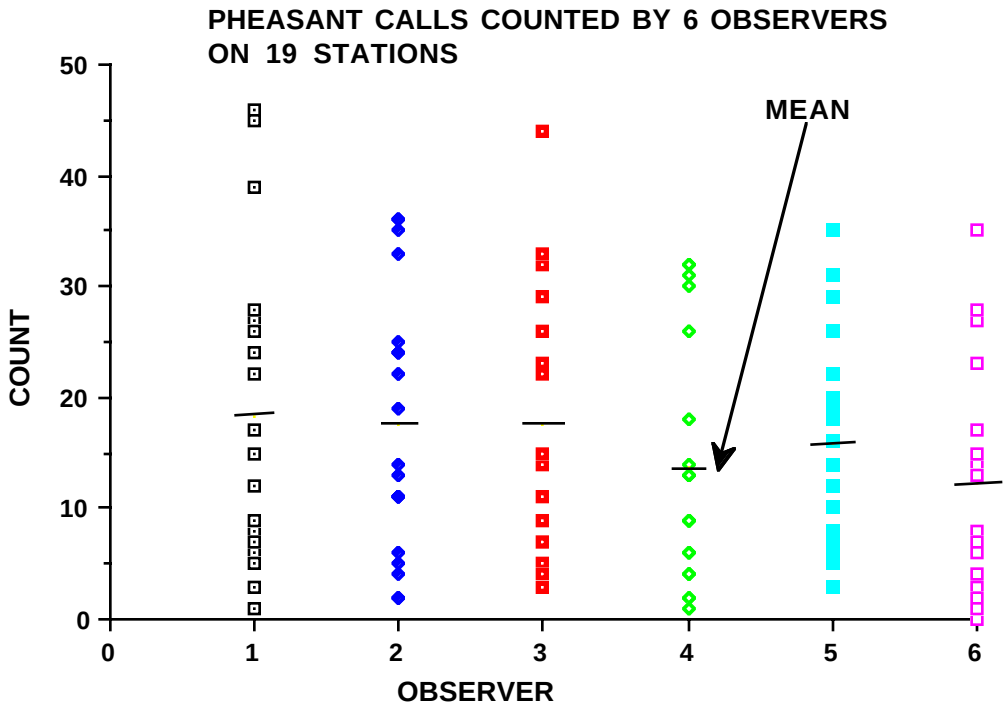


Fig. 6.1 Counts of calling (crowing) pheasants at 19 stations counted simultaneously by each of 6 observers.

The short horizontal lines in Fig. 6.1 mark the mean counts for each of the 6 observers. If it is assumed that the data all come from the same population or process, then the apparent differences in means arise as a matter of chance. Then any particular cluster of points will occupy roughly the same position as any other cluster. On the other hand, if at least some of the populations (or processes) do have quite different means, then the clusters of plotted points will not occupy quite the same positions. Three ways in which the clusters of points can differ are: (1) one or more clusters are shifted up or down from the others (a "scale" or "location" difference), (2) the spread of the individual clusters may differ, and (3) the shape of the clusters may differ. Sample variances provide a measure of the spread of the data, being calculated for the n_i observations from each observer as:

$$s_i^2 = \frac{\Sigma(y_i - \bar{y}_i)^2}{n_i - 1} \qquad (6.1)$$

so that data with a wide spread (scatter) of points will have a large variance. If the clusters have the same shape (this can't be reliably checked without very large samples) and spread (same variance), then a simple shift up or down scale can be detected by comparing the variability of individual clusters with that of the whole set of data. If the clusters are shifted well apart, obviously an overall variance will considerably exceed that of the individual clusters. Offhand, it doesn't look as though the several sets of pheasant data differ much. Another example, based on natural logarithms of counts of "signs"

(mounds and dens) of pocket gophers at different locations appears in Fig. 6.2, which does seem to suggest real differences between sites.

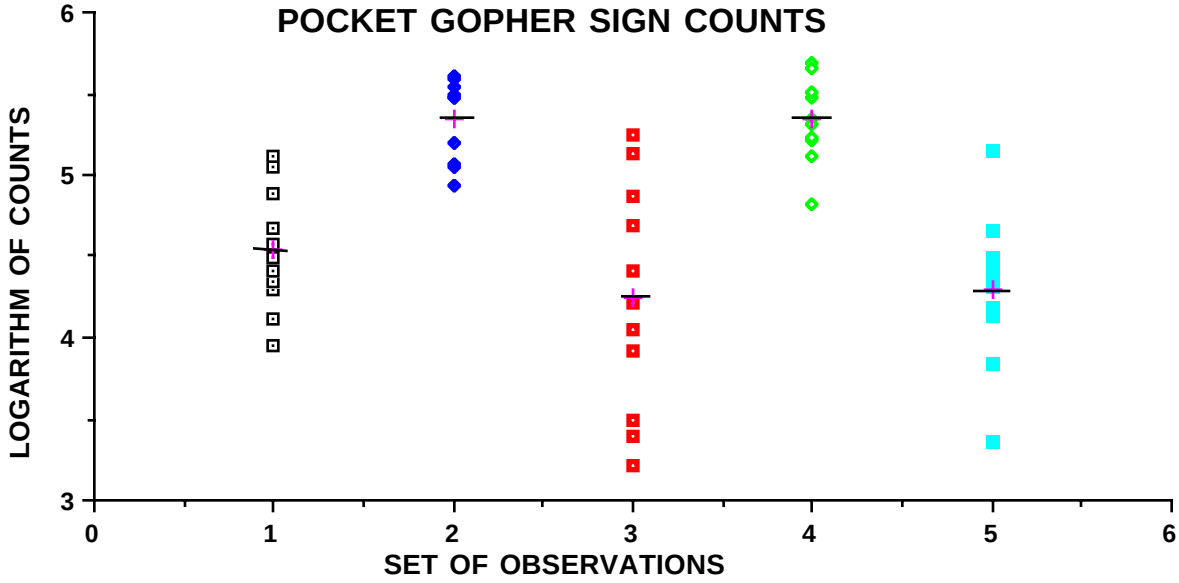


Fig. 6.2 Logarithms of counts of pocket-gopher signs at different locations and/or years (Reid, Hansen, and Ward, Jour. Wildl. Manage. 30:330,1966).

A comparison between an overall variance and that of individual comparisons can be constructed by examining the sum of squares making up the overall variance. That is, let y_{ij} be the j^{th} observation in the i^{th} column of tables of data and n_i be the number of observations in that column. Denote the overall mean by $\bar{y}$ and a column mean by $\bar{y}_i$. Then the overall variance is written as:

$$\frac{\sum_i \sum_j (y_{ij} - \bar{y})^2}{\sum_i n_i - 1} \qquad (6.2)$$

Considering only the numerator of eq. (6.2) (the sum of squares) for the present, we can rewrite it as:

$$\sum_i \sum_j (y_{ij} - \bar{y})^2 = \sum_i \sum_j [(y_{ij} - \bar{y}_i) + (\bar{y}_i - \bar{y})]^2 \qquad (6.3)$$

$$= \sum_i \sum_j (y_{ij} - \bar{y}_i)^2 + \sum_i \sum_j (\bar{y}_i - \bar{y})^2$$

$$\sum_i \sum_j (y_{ij} - \bar{y})^2 = \sum_i \sum_j (y_{ij} - \bar{y}_i)^2 + \sum_i n_i (\bar{y}_i - \bar{y})^2 \qquad (6.4)$$

TOTAL S.S. = WITHIN S.S. + BETWEEN S.S.

This results because a little algebra shows that the cross-product term vanishes (students should do the algebra for themselves). The first of the two resulting terms is just the sum of the components that would be used to calculate a separate variance for each column and is thus denoted the "within"

(within columns) sum of squares. The second component represents the variability "between" columns. These quantities are usually displayed in an Analysis of Variance (ANOVA) table (let $\Sigma n_i = n$):

Table 6.2 Analysis of Variance for a one-way design.

Source	Sum of squares	Degrees of freedom	Mean squares
Between groups	$SSb = \sum_i n_i(\bar{y}_i - \bar{y})^2$	k - 1	$\frac{SSb}{k-1}$
Within groups	$SSw = \sum_i \sum_j (y_{ij} - \bar{y}_i)^2$	n - k	$\frac{SSw}{n-k}$
Total	$= \sum_i \sum_j (y_{ij} - \bar{y})^2$	n - 1	

The "mean squares" (MS) are estimates of variances, and under the hypothesis of no difference between the populations (processes) represented by the columns of the Figures above, these estimates should be equal. Arriving at the divisors (degrees of freedom) can be remembered by the following devices: (1) there are k means being considered in the "between" groups so the usual practice for estimating a variance prevails, i.e., divide by k-1, (2) within each group a variance would be estimated by Eq. (6.1). A logical way to pool these within-group variances is to weight by the degrees of freedom, i.e., calculate:

$$\frac{\sum_i (n_i - 1)s_i^2}{\sum_i (n_i - 1)} \tag{6.5}$$

which gives the between-groups value used above (Table 6.2).

Whether the two variance estimates are significantly different or not is tested by the "F-ratio", which is:

$$F = \frac{SS_b/(k-1)}{SS_w/(n-k)}$$

Values denoting significant deviations are widely tabulated in textbooks in statistics and are now printed out by the various computer programs used to calculate ANOVAS. The advent of such computer programs has made it very easy to do the calculations. The serious disadvantage of these "canned" programs is that virtually anyone can calculate complex analyses without having any real idea what the results mean. Students thus need to actually work out the calculations for the above examples so as to understand how they are carried out. This is easy to do on a spreadsheet, such as EXCEL.

Inasmuch as EXCEL will conduct one-way ANOVAS, we can first use that function (Anova: single factor) and then calculate the sums of squares directly

on a separate spreadsheet, as a way to understand what's going on. Thus the EXCEL one-way program produces a listing of sample sizes, sums, averages and variances for each column in the table of pheasant call-count data, followed by an ANOVA table of Sums of Squares, degrees of freedom, mean squares, F value, P-value (probability of significant difference between groups, and "F-crit" (the significant value of F at the $\alpha = 0.05$ level).

To check these results directly, one needs only to insert two columns between each of the existing columns of data, calculate column means ($\bar{y}_i$)for the data and the overall mean ($\bar{y}$), and use these to calculate "within" and "total" sums of squares and add them up to get the values produced by the program. The "between" sum of squares is calculated directly from the definition given in Table 6.2 above using column means and overall mean.

6.2 Two-way analysis of variance

One-way ANOVA usually does not involve much in the way of a study design. The comparisons are likely to be obvious, and the only complication that may arise is if it is desired to compare subgroups of the k sets of observations. We will return to such comparisons later on. The "higher-order" forms of ANOVA are more versatile and thus more powerful. More planning is thus involved, and we need to distinguish between the various possible approaches. The simplest of the more complex ANOVA's is the two-way analysis without replications. As the name suggests, it is based on a two-way table. There are k sets of data, each appearing in r rows, so that there are rk observations. The pheasant call-count data provide an example, where we now consider the rows (stations) as a factor in the analysis. This is done by calculating a row sum of squares, and incorporating it in the ANOVA table. It is worthwhile to depict the data as an table of x_{ij} with k columns and r rows as follows (some authors use r rows and c columns; others a rows and b columns -- notation is not consistent in statistics books). It is useful to border the table with row and column means. The dot notation (e.g, $\bar{x}_{1.}$) is used to signify that the average is taken over a row or a column ($\bar{x}_{.1}$). A double dot notation ($\bar{x}_{..}$) is used to designate the overall means (sometimes this appears with two bars over x). **Note that we have switched from y_{ij} to x_{ij}.** Both notations are common; it is worthwhile to use x_{ij} from now on because y_{ij} will be used as the "independent" variable in regression analysis later on.

		COLUMN MAIN EFFECT					
		1	2	3	i	k	
	1	x_{11}	x_{21}	x_{31} ...	x_{j1}	.... x_{k1}	$\bar{x}_{\cdot 1}$
	2	x_{12}	x_{22}	x_{32}	x_{j2}	x_{k2}	$\bar{x}_{\cdot 2}$
ROW EFFECT	3	x_{13}	x_{23}	x_{33}	x_{j3}	x_{k3}	$\bar{x}_{\cdot 3}$
		.					
		.					
		.					
	j	x_{1j}	x_{2j}	x_{3j}	x_{ij}	x_{ki}	$\bar{x}_{\cdot i}$
		.					
		.					
		.					
	r	x_{1r}	x_{2r}	x_{3r}	x_{ir}	x_{kr}	$\bar{x}_{\cdot r}$
		$x_{1\cdot}$	$x_{2\cdot}$	$x_{3\cdot}$	$x_{i\cdot}$	$x_{k\cdot}$	$x_{\cdot\cdot}$

The sums of squares (S.S.) are obtained in the same way as in the previous example, that is, we expand the Total S.S. to form the other sums of squares:

Total S.S. - Columns S.S. - Rows S.S. = Residuals (Error) S.S.

$$\sum_{i=1}^{k}\sum_{j=1}^{r}(x_{ij}-\bar{x}_{\bullet\bullet})^2 - r\sum_{j=1}^{r}(x_{i\bullet}-\bar{x}_{\bullet\bullet})^2 - k\sum_{j=1}^{r}(\bar{x}_{\bullet j}-\bar{x}_{\bullet\bullet})^2 = \sum_{i=1}^{k}\sum_{j=1}^{r}(x_{ij}-\bar{x}_{i\bullet}-x_{\bullet j}+\bar{x}_{\bullet\bullet})^2 \qquad (6.6)$$

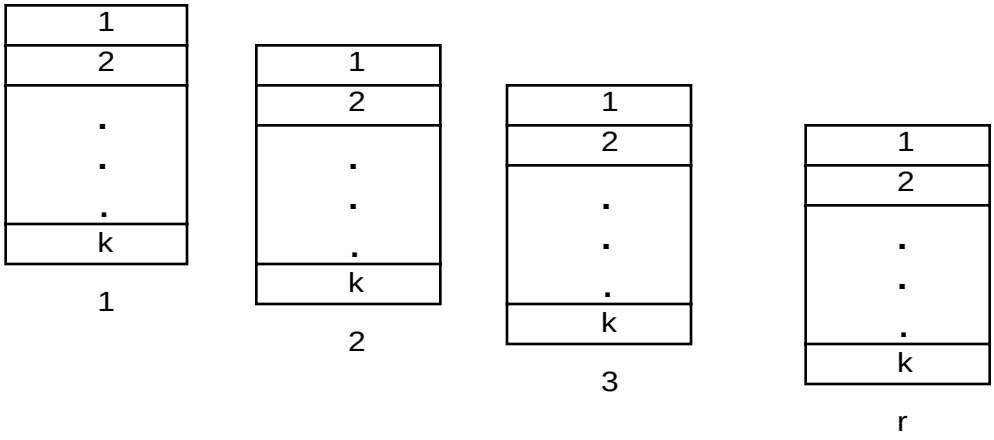
These results are calculated by EXCEL as 2-way ANOVA without replication. The program produces tables of row and column means and variances along with an ANOVA table.

6.3 Randomized blocks designs

The two-way program is listed in EXCEL as being "without replication". However, this is not necessarily true, as the row effects can indeed represent replications. Such an arrangement results from a randomized blocks design. These designs are widely applicable. Suppose we have k treatments to study, and can arrange to test them in r "blocks", where each block is comprised of k units that are relatively uniform in nature. For example, we might want to evaluate the effectiveness of k drugs on weight gain in rats. We might thus obtain r litters of k rats each, and give the different drugs to each of the k rats of each litter (choosing individual rats out of a given litter at random to receive one of the k drugs), and maintain the individual litters together under uniform conditions. The trick is to keep the blocks as uniform as possible so as to minimize "within block" variability so that most of the variance within a block results from the treatments. The method was developed in agricultural experimentation where the blocks are usually plots of ground selected for

their uniformity. Each plot is made up of k subplots, to which some set of, say, fertilizers, is applied. Fig. 6.3 shows how randomized blocks designs are laid out in plots. Note that the blocks may be separated by some distance, being selected for the uniformity of material within a block, which can reduce the "error" M.S. considerably.

The randomized blocks design can be a powerful and efficient approach. Note that the individual units in the blocks serve as true replicates so that the randomized blocks design does have replication. In our pheasant example, the stations are not replicates, so the ANOVA there is "without replication". However, the two cases (with and without replication) use the same calculations. The difference lies in the experimental design -- randomized blocks may be far more efficient in assessing differences. Much of the efficiency depends on the investigator's knowledge of the experimental material -- there is an element of "art" in picking blocks. In long-term studies one can sometimes take advantage of previous year's data to see how uniform the blocks are. Also, "uniformity" studies can be run to measure the variability within blocks. In these studies the same "treatment" (usually no treatment) is applied to all plots, and the ANOVA run to measure variability within and between blocks. One would, of course, like to have a very small "within" mean square, and can tolerate a large "between" blocks M.S.



A randomized block study design. T here are k units, assigned to a lo cation at random within each block, and r blocks in all. Every treatment ap pears in each block (randomly assigned t o a position).

Fig. 6.3 Randomized blocks layout. The blocks (often plots in agricultural studies) are laid out to be as uniform as possible within individual plots.

6.4 Two-way analysis of variance with replication

The two-way analysis of variance with replication normally appears with replicates "within cells" in a table of data. We thus need to consider observations with three subscripts, x_{ijk}, as shown in the table below, which has 3 replicates per cell. In general, we may have m replicates per cell where $m \geq 2$, and thus rkm observations in the entire table.

The calculations for S. S. in the ANOVA table now become somewhat more complicated, but take on a general form that can be followed in even more complex cases. The residual S. S. (error term) is always calculated from the replicates within cells, i.e.,

$$\text{Residual S.S.} = \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij\bullet})^2 \qquad (6.7)$$

where $\bar{x}_{ij\cdot}$ is the cell mean (these are not shown in the table below as they are an average of the m observations in the cell; 3 observations in the table above). An easy way to remember how to calculate residual (error) mean squares when there is replication, is to note that the units within a given cell all get identical treatments, and thus furnish the best estimate of the underlying variability. Hence the error mean square estimates the underlying variance of the experimental units. Any other variance estimate (mean square) may be inflated by treatment effects.

COLUMN MAIN EFFECT

		1	2	3	i	k	
		x_{111}	x_{211}	x_{311} ...	x_{j11}....	x_{k11}	
	1	x_{112}	x_{212}	x_{312} ...	x_{j12} ...	x_{k12}	$\bar{x}_{\cdot 1\cdot}$
		x_{113}	x_{213}	x_{313} ...	x_{j13} ...	x_{k13}	
		x_{121}	x_{221}	x_{321}	x_{i21}	x_{k21}	
	2	x_{122}	x_{222}	x_{322}	x_{i22}	x_{k22}	$\bar{x}_{\cdot 2\cdot}$
		x_{123}	x_{223}	x_{323}	x_{i23}	x_{k23}	
	•	•	•	•	•	•	
	•	•	•	•	•	•	
	•	•	•	•	•	•	
		x_{1j1}	x_{2j1}	x_{3j1}	x_{ij1}	x_{kj1}	
ROW EFFECT	j	x_{1j2}	x_{2j2}	x_{3j2}	x_{ij2}	x_{kj2}	$\bar{x}_{\cdot j\cdot}$
		x_{1j3}	x_{2j3}	x_{3j3}	x_{ij3}	x_{kj3}	
	•	•	•	•	•	•	
	•	•	•	•	•	•	
	•	•	•	•	•	•	
		x_{1r1}	x_{2r1}	x_{3r1}	x_{ir1}	x_{kr1}	
	r	x_{1r2}	x_{2r2}	x_{3r2}	x_{ir2}	x_{kr2}	$\bar{x}_{\cdot r\cdot}$
		x_{1r3}	x_{2r3}	x_{3r3}	x_{ir3}	x_{kr3}	
		$x_{1\cdot\cdot}$	$x_{2\cdot\cdot}$	$x_{3\cdot\cdot}$	$x_{i\cdot\cdot}$	$x_{k\cdot\cdot}$	$x_{\cdot\cdot\cdot}$

Another very general S. S. is the Total sum of squares, calculated from the individual observations as before:

$$\text{Total S.S.} = \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{\bullet\bullet\bullet})^2 \qquad (6.8)$$

with $\bar{x}_{\bullet\bullet\bullet}$ ... the overall mean.

The third general S. S. is the Treatment Sum of Squares, calculated from the cell means

$$\text{Treatment S.S.} = m\sum_i \sum_k (x_{ij\bullet} - \bar{x}_{\bullet\bullet\bullet})^2 \qquad (6.9)$$

If any of the treatments are effective, the treatments mean square will be inflated. Of course, we want to be able to break this overall S.S. down into row and column S. S. . As before, we do this with row and column means:

$$\text{Row S. S.} = mr\sum_{i=1}^{k} (\bar{x}_{i\bullet\bullet} - \bar{x}_{\bullet\bullet\bullet})^2 \qquad (6.10)$$

$$\text{Column S.S.} = mk\sum_{j=1}^{r} (x_{\bullet j\bullet} - \bar{x}_{\bullet\bullet\bullet})^2 \qquad (6.11)$$

The two-way analysis with replication contains a new S. S., the interaction Sum of Squares. This is often calculated as Treatment S. S. - Row S. S. - Column S. S. but a direct calculation from the means is

$$\text{Interaction S.S.} = m\sum_{i=1}^{k} \sum_{j=1}^{r} (\bar{x}_{ij\bullet} - \bar{x}_{i\bullet\bullet} - \bar{x}_{\bullet j\bullet} - \bar{x}_{\bullet\bullet\bullet})^2 \qquad (6.12)$$

The Total S.S. breaks down into Treatments and Error, and the Treatments S.S. contains Rows, Columns and Interaction S.S. Textbooks usually show the ANOVA table in this form, but EXCEL ignores Treatments, producing only Total, Rows (labelled Samples for unknown reasons), Columns, and Interaction.

In the final ANOVA, the F-test of significance of the interaction mean square (MS_{Inter}/MS_{Error}) is very important in deciding what can be said about the main effects. This is because a significant interaction mean square suggests that the row and column main effects are somehow correlated, i.e., they "interact". If this is the case, then one cannot discuss the two sets of main effects (row and column) separately, making interpretation of the experiment much more difficult.

Note the similarity of the equation for Interaction S. S. to that for the Residuals (error) S. S. for the 2-way ANOVA without replication. This suggests that the error term in that case is really an interaction term, making it evident that we need to have replications to assess interaction (there is a test, Tukey's test, for a particular form of interaction in a 2-way ANOVA based on one observation per cell. It appears in many statistics texts (e.g., Snedecor and Cochran , Statistical Methods, Iowa State University Press, Ames, Iowa, 6th Edition, 1967).

The pheasant count data of Table 1 were not replicated in a strict sense, which would require repeated counts at the same station by the same observers. One of the reasons this is not done is that calling activity drops off quite sharply after the early morning hours. It is also obvious from Table 1 that there is a gradient over distance. High counts are obtained in the best pheasant habitat, and this route apparently went into marginal habitat beyond Station 10. Inasmuch as the stations are reasonably close together (usually 1 mile apart to keep from counting the same calling individuals a second time), it isn't too much of a stretch of technique to regard adjacent stations as "replicates". To do this, we drop the counts on Station 19, and use successive pairs as replicates (thus counts on Stations 1 and 2 are called replicates, while Stations 3 & 4 are also replicates, etc., giving r = 9, while k= 6 for rkm= 9(6)2 = 108). The analysis of variance table is as follows.

Table 6.3 Analysis of variance of Pheasant count data with m=2.

Source	SS	df	MS	F
Rows	11189.91	8	1398.74	52.24
Columns	568.41	5	113.68	4.25
Interaction	605.76	40	15.14	0.57
Within	1446.00	54	26.78	
Total	13810.07			

6.5 Assumptions for the analysis of variance

The discussion thus far has focussed on the mechanics of the analysis of variance, being mainly concerned with developing Sums of Squares for 3 models: (1) one-way ANOVA, (2) 2-way ANOVA without replication (and the special case of randomized blocks designs), and (3) two-way ANOVA with replication. The ANOVA tables present the S. S., their associated Mean Squares, and the ratios of Mean-Squares (F-ratios). As previously noted, the tests of significance (F-tests) depend on the assumptions underlying ANOVA, but the underlying framework - the Sums of Squares, along with the Mean Squares and F-ratios, can be calculated for any set of numbers. No assumptions are required. We thus have a mechanical analysis that says something about variability introduced by treatments, without assumptions about the underlying data.

To consider the assumptions required for tests of significance, we write a model for the observations for a 2-way ANOVA with replications:

$$x_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ij} \qquad (6.13)$$

Here, μ represents an overall mean value, α_i and β_j are the main effects (column and row effects), γ_{ij} represents the interaction between the two main effects, and ε_{ij} is the error term. This latter term (ε_{ij}) is assumed to have an "expected value" of zero. That is, when it is averaged over a large data set it should equal zero. Usually it is assumed that $\Sigma\alpha_i = 0$ and $\Sigma\beta_j = 0$. We thus have the

x_{ij} made up of an overall mean value (μ) plus an effect for its row (α_i) and its column (β_j). As we noted earlier, an "interaction" is an effect that makes adjacent observations tend to be correlated. When there are no interactions ($\gamma_{ij} = 0$) then the expected value of x_{ij} (effectively x_{ij} averaged over very large samples) can be written as $E(x_{ij}) = \mu + \alpha_i + \beta_j$ and we say that the model is additive. Such analyses are far easier to understand and interpret than are those where interactions are present (γ_{ij} not equal to zero).

It is important to recall that we want to test several hypotheses that state that the main effects and interaction are zero, and the assumptions become important in assuring validity of the F-tests.

Assuming additivity, we can use the reduced model:

$$x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \qquad (6.14)$$

Two major assumptions underlying tests of significance in ANOVA are:

(1) the ε_{ij} are independent, i.e., uncorrelated.

(2) the ε_{ij} are from a normal distribution with mean zero and variance σ^2. The normal distribution is a symmetrical, bell-shaped curve, with its "spread" (variance) measured by σ^2 [Eq.(1.3)].

Consider a two-way ANOVA with many replications per cell (and no interactions). The mean value in any cell should be approximately

$$\bar{x}_{ij.} = \mu + \alpha_i + \beta_j$$

and the x_{ijk} in this cell should have the same variance, σ^2. Any two cells should have the same variance, σ^2. This is often described as homoscedasticity, which simply means equal variances.

The assumptions for ANOVA can be listed as:

(1) additivity ($\gamma_{ij} = 0$).
(2) independence of the ε_{ij}
(3) ε_{ij} normally distributed with mean zero and variance σ^2.

Sometimes (3) is split into 2 assumptions:
(3) ε_{ij} normally distributed with mean zero
(4) homoscedasticity - variances in replicates are all equal to σ^2.

In most applications of ANOVA there simply are not enough replicates within cells to test these assumptions. Given quite large samples in the cells (say 20-30 replicates per cell) it is worth comparing variances. Testing for

normality takes larger samples. Some authors recommend Bartlett's test, but Scheffe (The Analysis of Variance, J. Wiley and Sons, 1959, p. 83) points out that it "is extremely sensitive to nonnormality", and recommends that a preliminary test of homogeneity of variances not be made.

Ecological data often come as counts of some kind, and these tend not to be normally distributed, often having a skewed frequency distribution -- a long "tail" of less frequent observations on one side or the other of the bulk of the observations. Such data can be brought into closer approximation to normality by a transformation. Two of the most commonly used transformations are the square root transformation $(x_{ij})^{0.5}$, and the logarithmic transformation, $\log_e(x_{ij})$. It often turns out that standard deviations of ecological data tend to be proportional to the mean values (coefficient of variation, $s/\bar{x}$ = approximately a constant). The logarithmic transformation tends to "normalize" such data and to make variances more nearly equal on the transformed (i.e., logarithmic scale).

Testing the need for or the effects of a transformation is often recommended, but it is risky to let such tests govern a decision to use or not use a transformation.

It is worthwhile to simulate data based on the assumptions for ANOVA. We start with Eq.(6.14); no interactions, and produce a table of main effects (using α_i and β_j such that they sum to zero) to which we add μ (taken as 5 here). A table of random normal deviates can be produced using EXCEL (used here as N(0,1), i.e. normal with zero mean and unit variance). These are added to the main effects table giving a set of simulated data (3 replicates per cell were used; they have the same main effect value, but different random draws were used to add ε_{ij}). The table of data follows, along with the ANOVA table.

DATA FOR ANOVA

	1	2	3	4
1	6.10	5.89	4.98	5.67
	7.18	4.80	5.51	5.70
	5.73	6.71	4.89	6.80
2	6.33	6.17	6.75	2.91
	6.48	5.89	2.54	4.13
	6.77	4.42	5.56	5.23
3	6.57	5.36	5.93	3.71
	6.93	4.93	3.08	4.80
	5.68	5.42	2.81	5.31
4	6.49	4.14	5.80	2.65
	3.39	4.74	5.26	5.82
	4.49	6.39	4.74	3.89
5	4.18	5.97	5.14	2.61
	6.55	5.18	4.46	3.11

	4.10	3.99	3.93	3.57
6	5.69	4.40	4.20	4.25
	4.15	3.63	4.18	4.78
	3.97	2.88	3.75	4.49

Source	SS	df	MS	F	P-value	F crit
Sample	21.20	5.00	4.24	4.08	0.00	2.41
Columns	14.67	3.00	4.89	4.71	0.01	2.80
Interaction	18.29	15.00	1.22	1.17	0.32	1.88
Within	49.86	48	1.04			
Total	104.02	71				

We can repeat the above exercise with interactions. In this case, the γ_{ij} were taken as a fractional power of the product denoting row and column positions of a main effect entry, $\gamma_{ij} = (xy)^{0.7}$. Again a random normal deviate is added to give eq. (6.13). The table of "data" and ANOVA table follow.

		1	2	3	4
	1	7.10	7.51	7.14	8.30
DATA		8.18	6.42	7.67	8.34
FOR		6.73	8.33	7.05	9.43
ANOVA	2	7.96	8.81	10.25	7.20
		8.10	8.53	6.04	8.42
		8.40	7.06	9.07	9.51
	3	8.73	8.87	10.59	9.40
		9.09	8.44	7.73	10.50
		7.84	8.93	7.47	11.00
	4	9.13	8.43	11.49	9.62
		6.03	9.03	10.96	12.79
		7.13	10.67	10.44	10.85
	5	7.27	10.98	11.79	10.75
		9.63	10.19	11.12	11.25
		7.19	9.00	10.58	11.71
	6	9.19	10.09	11.77	13.50
		7.65	9.33	11.75	14.03
		7.47	8.57	11.31	13.74

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Sample	78.15	5	15.63	15.05	0.0000
Columns	69.00	3	23.00	22.14	0.0000
Interaction	44.39	15	2.96	2.85	0.0030
Within	49.86	48	1.04		
Total	241.40	71			

As noted previously, some authors recommend testing the deviations (given in Eq.(6.12)) for normality. With the numbers of replicates usually

available, this is not sensible advice. A plot (Fig. 6.4) of all 72 normal deviates used in the simulations was generated from a normal distribution, but is not too reassuring in terms of the assumption of normally distributed errors.

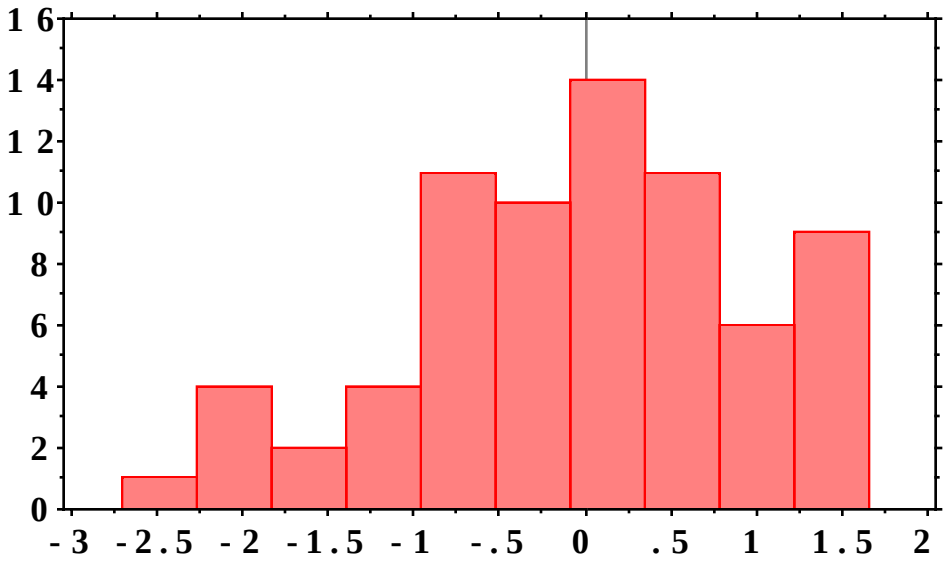


Fig. 6.4. Frequency distribution of 72 deviations from data used in simulations.

6.6 Comparisons in ANOVA

In a brief account like the present one, it is not possible to cover more than a fraction of the features of the Analysis of Variance. The book by Scheffe is a classic account, and should be examined for more details. It is, however, couched in the language of matrix algebra. Another good account is that of Snedecor and Cochran, Statistical Methods, the Iowa State University Press, Ames, Iowa. It has gone through at least 8 editions, and is another classic text. Important topics that we have not covered are those of comparisons or contrasts. In most experimental work, the main interest will be in certain comparisons (e.g. Exercise 6.12 on weight gains in rats). In the pheasant data there were 2 observers with experience, so a comparison between experienced and inexperienced observers is of considerable importance. The pocket-gopher data (Exercise 6.5) was collected at different locations and over different years. One would thus emphasize those comparisons. In Exercise 6.8 there is a "control" plot ("check" treatment) which would normally be compared with all other plots. Snedecor and Cochran give good descriptions of how to sort out such contrasts. A short account of two approaches follows.

When comparisons are planned in advance, a t-test can be used to test these specific comparisons for significance. The test depends on computing a linear combination of the observed means:

$$L = \lambda_1 \bar{x}_1 + \lambda_2 \bar{x}_2 + \ \ldots \ + \ \lambda_k \bar{x}_k \qquad (6.15)$$

with the λ_i constants adding to zero, i.e. $\Sigma \lambda_i = 0$. The standard error of L is estimated as:

$$S.E.(L) = s_L = \sqrt{\Sigma \lambda_i^2}\,\frac{s}{\sqrt{n}} \qquad (6.16)$$

The d.f. for the estimated standard error of L are those used to estimate s, and n is the number of observations used to compute each mean, $\bar{x}_i$. Scheffe describes comparisons such as L as contrasts. The t-test for comparisons planned in advance is:

$$t = \frac{L}{s_L} \qquad (6.17)$$

The λ_i are dictated by the comparison desired. If, for example, 2 means, say $\bar{x}_1$, and $\bar{x}_3$, are compared to a third one, $\bar{x}_2$ then $L = \frac{\bar{x}_1}{2} + \frac{\bar{x}_3}{2} - \bar{x}_2$ with the λ_i being 1/2,1/2 and -1, and thus adding to zero. If there are additional means in the overall analysis that are not to be involved in the comparison, then the λ_i for those means are assumed to be equal to zero. The simplest comparison is that for 2 means, with the comparison being $\bar{x}_1 - \bar{x}_2$, so that $\lambda_1 = 1$ and $\lambda_2 = -1$, and
$s_L = 2^{1/2}s/n^{1/2}$ so that:

$$t = \frac{n^{1/2}(\bar{x}_1 - \bar{x}_2)}{2^{1/2}s}$$

with n being the number in each group and s is obtained from the error M.S.

A test due to Scheffe provides a general method for finding significant differences among a full set of means without designating these comparisons in advance of conducting the experiment. The price paid is less sensitivity (broader confidence limits). It uses the same set-up as above, but declares L/s_L significant only if it exceeds $[(k-1)F_{05}]^{1/2}$ where F_{05} is the 5% level of the F-distribution for k-1 and n-k degrees of freedom when we are considering a one-way analysis. The test can be used in more complex ANOVAs using (k-1) and the d.f. associated with the error mean square. Scheffe's test also reduces to the t-test when k = 2. The test should not be used if the F-test in an ANOVA is not significant, as there are then no significant contrasts in the data. It is important to understand that the S-method can be used to check all significant contrasts in the means, while preserving the chosen α level.

We illustrate the two procedures by using data simulated as in Section 6.5 where the simulations were used to study the assumptions for ANOVA. In this example the column main effects have been changed. The "data" are as follows:

	1	2	3	4
1	5.90	5.39	5.48	5.87
	6.98	4.30	6.01	5.90
	5.53	6.21	5.39	7.00
2	6.13	5.67	7.25	3.11
	6.28	5.39	3.04	4.33
	6.57	3.92	6.06	5.43
3	6.37	4.86	6.43	3.91
	6.73	4.43	3.58	5.00
	5.48	4.92	3.31	5.51
4	6.29	3.64	6.30	2.85
	3.19	4.24	5.76	6.02
	4.29	5.89	5.24	4.09
5	3.98	5.47	5.64	2.81
	6.35	4.68	4.96	3.31
	3.90	3.49	4.43	3.77
6	5.49	3.90	4.70	4.45
	3.95	3.13	4.68	4.98
	3.77	2.38	4.25	4.69
Means	5.3995	4.5498	5.1398	4.6118

The ANOVA (two-way with replications) is:

Source	*SS*	*df*	*MS*	*F*	*P-value*
Sample	21.1987	5	4.2397	4.0818	0.0036
Columns	9.1840	3	3.0613	2.9473	0.0421
Interaction	18.2903	15	1.2194	1.1739	0.3233
Within	49.8576	48	1.0387		
Total	98.5306	71			

If we suppose the planned comparison was between means 1 and 3 against means 2 and 4, then:

$L = 0.25(5.3995) - 0.25(4.5498) + 0.25(5.1398) - 0.25(4.6118) = 0.3444$

and: $s_L = (\Sigma \lambda_i^2)^{1/2} \frac{s}{n^{1/2}} = 0.5[(1.0387)/(18)]^{1/2} = 0.122$

then: $t = \frac{L}{s_L} = \frac{0.3444}{0.122} = 2.82$ with 48 d.f.

From t-tables ($\alpha = 0.05$) we have $0.005 < P < 0.010$. EXCEL has a function that will compute the probability directly. Enter the statement = TDIST(t,d.f.,tails) where t is the calculated value (2.82 here), d.f. are 48, and "tails" is 2 for a 2-tailed test. This function yields P = 0.007. Quite possibly past experience would lead to a one-tailed test for a planned comparison.

We can illustrate Scheffe's S-method for the same comparison. The t-value remains the same (2.82) but we use:

$$[(k-1)F_{05}]^{1/2} = [3(2.80]^{1/2} = 2.90$$

as the criterion for significance at the 5% level, where 2.80 is the tabulated F-value at $\alpha = 0.05$ with 3 and 48 d.f. Hence, the test is close to the 5% level of significance. We can go on and look for other significant contrasts as suggested by the data while still having $\alpha = 0.05$. This is definitely not the case for the first comparison tested above, which has to be selected in advance of the study. Because it is a general-purpose "data-snooping" tool, Scheffe (1959) suggested his test might be used with $\alpha = 0.10$, rather than the usual 0.05. EXCEL can be used to find the tabular F-value by using the function FINV(P, d.f.1, d.f.2) where P = 0.05 here and, d.f.1 = 3, and d.f.2 = 48. This function gives F_{05} = 2.798, and F_{10} = 2.201.

6.7 Type I and II errors and "power"

Most ecologists are used to the notion of Type I error, routinely conducting statistical tests, such as the t-test, at the 5% level of significance ($\alpha = 0.05$). They understand that such tests give a 0.05 probability, over the long run of many such tests, of erroneously claiming that the null hypothesis of "no effect" can be rejected when it is in fact true. Many do not seem to realize that there is another side to the issue, which is failing to find a significant difference when it exists (possibly because there were not enough samples or replications to detect an important difference. This is known as a type II error.

This issue of type II error can be discussed in terms of the "sensitivity" of a study, i.e., how small a change or difference will a study of a given size reliably detect? The statistician's answer is usually couched in terms of a power function or the "power of a test". Consider the likely points of view as to the impact on the environment of some new facility. There are usually two sides, those who construct and operate the facility and those with environmental regulatory authority. Suppose that both sides can agree on a study method that has well-known statistical properties and can be applied in the circumstances under consideration. Suppose that they further agree to make certain modifications in the facility and/or operational procedures if a field survey shows a specified degree of change has taken place (an "impact"). What remains is to decide how large a sample should be taken in the field study. But that depends on the amount of protection each party requires against errors damaging to their best interests.

These can be described as follows: (1) The people doing the construction and operation would rather not have the survey results indicate a significant change when the agreed-on degree of change really did not take place (Type I error). Just how strongly they voice objections will depend on the consequences of a determination of "change". If only minor modifications are then necessary they perhaps will agree that a 10% rate ($\alpha = 0.10$) of such "false positive" results is acceptable. However, if the changes require costly retrofitting and expensive operational modifications, they may well want to try to insist on Type I error rates of 1% or maybe even less.

(2) On the other side, the staff of the regulatory agency would not like to fail to recognize a significant impact when one does occur (Type II error). If small

samples are taken, the results almost always will come out not significantly different. Hence the regulators may be guided by rules that require an 80% chance of being sure to detect a real difference (of the magnitude agreed on) when the impact is not of minor environmental consequence. But if very substantial damage to an important resource may be involved, they may well argue for a 99% assurance. All too often, by default or lack of understanding the actual rate may be about 50%, much like settling the issue by flipping a coin and doing no field work.

To make any progress in the ensuing arguments a way of estimating sample sizes is needed. An easy solution is just to take a very large sample. Usually that is either too expensive or impractical (it may also result in environmental damage from the sampling process). A handy formula for approximating sample size for given Type I and II errors is given by Snedecor and Cochran and in the useful book by Cochran (Planning and Analysis of Observational Studies, W. G. Cochran 1983, J. Wiley and Sons). It can be written as:

$$n = \frac{2(z_\alpha + z_\beta)^2 \sigma^2}{\delta^2} = \frac{2(z_\alpha + z_\beta)^2}{\{\frac{\delta}{\sigma}\}^2} \qquad (6.18)$$

where z_α = normal deviate for Type I error, z_β = normal deviate for Type II error, σ^2 = variance (assumed the same in both data sets being compared), and δ = true (unknown) difference between two population means ($\mu_1 - \mu_2$) or two areas being studied, and n = the desired sample size for each population or area (thus 2n required). z_α is the familiar value used in confidence limits, i.e., $z_{05} = 1.96$, $z_{10} = 1.64$. Some values of z_β are:

Type II error (β)	Power ($1 - \beta$)	z_β
0.20	0.80	0.84
0.10	0.90	1.28
0.05	0.95	1.64
0.01	0.99	2.33

A major problem is that σ is always unknown, and must be either guessed at or estimated from a preliminary survey (etc.). Thus the right-hand side of eq. (6.18) is frequently used, i.e., one guesses at the ratio of the difference to be detected to σ. Suppose we take $z_\alpha = 1.96$ and $z_\beta = 1.28$ (power = 0.90). Then

$$n = \frac{2(1.96 + 1.28)^2}{\{\frac{\delta}{\sigma}\}^2} = \frac{21}{\{\frac{\delta}{\sigma}\}^2}$$

Consequently, if we suppose that the true difference is one-half of σ, n = 84, while n = 21 if we assume $\sigma = \delta$. Clearly, if we assume a small difference is to be detected, the sample size required may be huge. Using a small sample without

looking at power of the test amounts to operating in ignorance (but still happens a lot).

6.8 Other aspects of ANOVA

We did not go beyond two-way tables with replications. Efficient designs will use more factors in order to get the most information per dollar spent on experimentation. Again Snedecor and Cochran provide good discussions. There are text-books devoted entirely to the design and analysis of experiments, and a bewildering array of prospects. We also did not investigate what are called unbalanced analyses. These are typically two-way analyses with replications where the same number of replications per cell is not present. Sometimes a study is planned with m replicates per cell but some are destroyed or, in the case of experiments with animals, die unexpectedly, etc. In other cases, it may not be possible to get m replicates per cell. Analyses of unbalanced designs can be complex and, in some cases, controversial. It may be noted that the pocket-gopher example is unbalanced, but this is not a problem in one-way analyses.

The models described here are fixed-effects models, where interest is solely in the set of main effects studied in the experiment. Very often we have to consider random-effects models where the effect studied is regarded as a sample from some large population of effects. The analysis then takes a different form. Probably most practical work can be described by mixed-effects models, where one set of factors is fixed and the remainder random. The great advantage of the fixed-effects model is that the analysis of variance is quite "robust" in such cases, i.e., non-normality is not as serious a concern as in the random-effects models, where we assume sampling from a random normal distribution, and depend much more on that assumption for tests of significance. We remarked that significant interactions pose problems of interpretation, but did not note that it the may be necessary to use the interaction term as the denominator of F-tests.

The prominence of ANOVA in ecological studies is a bit puzzling. A quick review of a major ecological journal a few years back showed that ANOVA was then the dominant statistical technique used in that publication. However, I suspect that many of the cases really stem from editorial and reviewer insistence on statistical testing. The mere fact that some “significant” result was obtained doesn’t really provide much information about the process being studied. Hence I suggest that students use the data-snooping quality of the test due to Scheffe (Section 6.6) whenever possible as a tool for searching out the particular comparisons that really are significant in an analysis. As noted there, Scheffe proposed using a 10% significance level with the test, but that may not be very palatable to editors (who often will not realize that it is up to the investigator to choose the significance level, not an editor or a referee).

6.7 Exercises

6.1 Show that the cross-product terms in eq. (6.3) cancel out, resulting in eq.(6.4).

6.2 Calculate a one-way analysis of variance for the pheasant data using the ANOVA program in EXCEL.

6.3 Using the group variances printed out in the EXCEL output for Exer. 6.2, calculate a value for Eq. (6.5) and locate the corresponding value in the one-way ANOVA table prepared in Exer. 6.2 (the calculations can be inserted to the right of the summary of means and variances in the one-way output tables).

6.4 Copy the data from Exer. 6.2 to a new spreadsheet in the same Workbook and calculate the S. S. for the one-way ANOVA directly from the definitions given in eq. (6.4).

6.5 Repeat the calculations for Exer. 6.2 and 6.4 on the pocket gopher data given below, but first convert the counts to natural logarithms [natural log = LN(number)].

Pocket gopher counts

	Black Mesa			Grand Mesa	
	1962	1963	1964	1963	1964
1	61	237	82	167	62
2	61	271	68	183	172
3	132	253	168	297	106
4	158	157	190	188	87
5	90	244	33	238	89
6	52	180	57	285	81
7	107	269	30	124	29
8	73	138	25	209	75
9	155	159	50	248	65
10	77	237	131	204	46
11	82		108		
12	97				
13	167				

6.6 Run the two-way ANOVA without replications on the pheasant data. Notice that the total sum of squares remains the same as the one-way analysis, and the between-groups S. S. is the same as that for columns in the new analysis. However, the F-tests are now significant, and it is worth considering why this should happen (look at the mean square for error, and compare it with the within-groups value of the one-way analysis. What is your explanation?

6.7 Calculate the S. S. for Exer. 6.6 directly.

6.8 Randomized blocks The data below are from a randomized block study reported by Snedecor and Cochran (Statistical Methods, 6th Ed., p. 300). Do a two-way analysis in EXCEL and compare the M. S. for replication with that for treatments. What do the F-tests suggest?

	Replicates				
Treatments	1	2	3	4	5
CHECK	8	10	12	13	11
ARASAN	2	6	7	11	5
SPERGON	4	10	9	8	10
SEMESAN	3	5	9	10	6
FERMATE	9	7	5	5	3

6.9 Calculate the ANOVA table for Exer. 6.8 directly. It shouldn't take long, and is first-rate practice in using EXCEL.

6.10 Calculate ANOVA for the pheasant data arranged as having 2 replicates as described in the text section on 2-way ANOVA.

6.11 Copy the data from Exer. 6.10 to a new spreadsheet in EXCEL and calculate the S. S. directly from the definitions of eq. (6.8) thru eq.(6.12). This is something of a chore, with the main difficulty being in keeping things straight (use good labelling). Use two spreadsheets. You will need 2 copies of the original counts (one for computing Total S.S., the other for computing error S. S.) and 2 copies of the cell means (to calculate Treatment and Interaction S. S.). If you label things carefully and use the proper multipliers (m, k, and r) you will get the same S. S. as in Exer. 6.10. Patience is necessary, as is accuracy. If you do all of the exercises, it should help in remembering how to use ANOVA.

6.12 Snedecor and Cochran (1967:p. 347) give the following data on gains in weight (grams) of rats fed on six diets. The columns are replicates (individual rats on the same treatment). Run an ANOVA and report the results. Note that a two-way ANOVA with replications is indicated.

	High level			Low level		
	Beef	Cereal	Pork	Beef	Cereal	Pork
1	73	98	94	90	107	49
2	102	74	79	76	95	82
3	118	56	96	90	97	73
4	104	111	98	64	80	86
5	81	95	102	86	98	81
6	107	88	102	51	74	97
7	100	82	108	72	74	106
8	87	77	91	90	67	70
9	117	86	120	95	89	61
10	111	92	105	78	58	82

6.13 Use the planned comparison method on the data of Exer. 6.12 to compare the 2 levels of protein (High and Low). You may also want to look at the breakdown given by Snedecor and Cochran for this example, as it uses orthogonal comparisons to break down the treatment S.S. into 5 individual comparisons. Apply Scheffe's S-method to the data. Discuss results.

6.14 Use the planned comparison method to compare the "check" (control) mean with the other treatment means of the data in Exer. 6.8. Make the same comparison with Scheffe's method. List his criterion for α =0.10 as well as for α =0.05. Discuss results.